SKA science data products and construction timeline

Shari Breen - Head of Science Operations

(and thanks to Sarrvesh Sridhar who contributed to these slides)

LHCOPN-LHCONE meeting

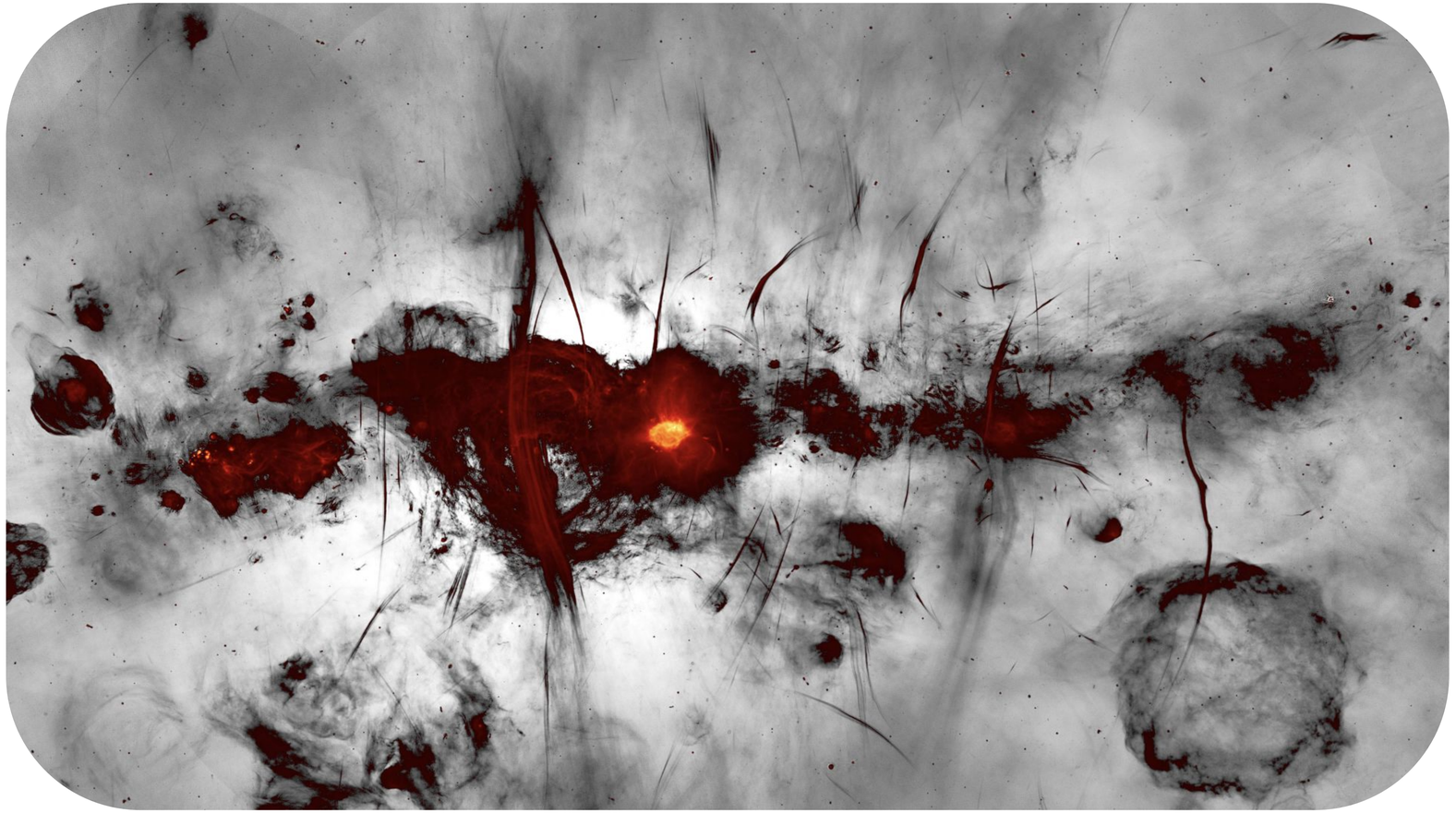
Sorry I am not there..

The SKAO operational model

SKA is a flexible science machine!

- SKA systems are hugely flexible
  - TWO telescopes to cover a frequency range between 50 MHz and 15.4 GHz
  - Each supports up to 16 subarrays (splitting the 512 stations and 197 dishes into smaller arrays)
  - Very flexible Correlator beam formers (CBFs) but ultimately resource limited
  - Both imaging and non-imaging modes
    - Broad-band continuum, Spectral/zoom, Pulsar and transient search (PSS), pulsar timing (PST), VLBI
  - Commensality supported (data, observing, multiplexed)

We will deliver data products!

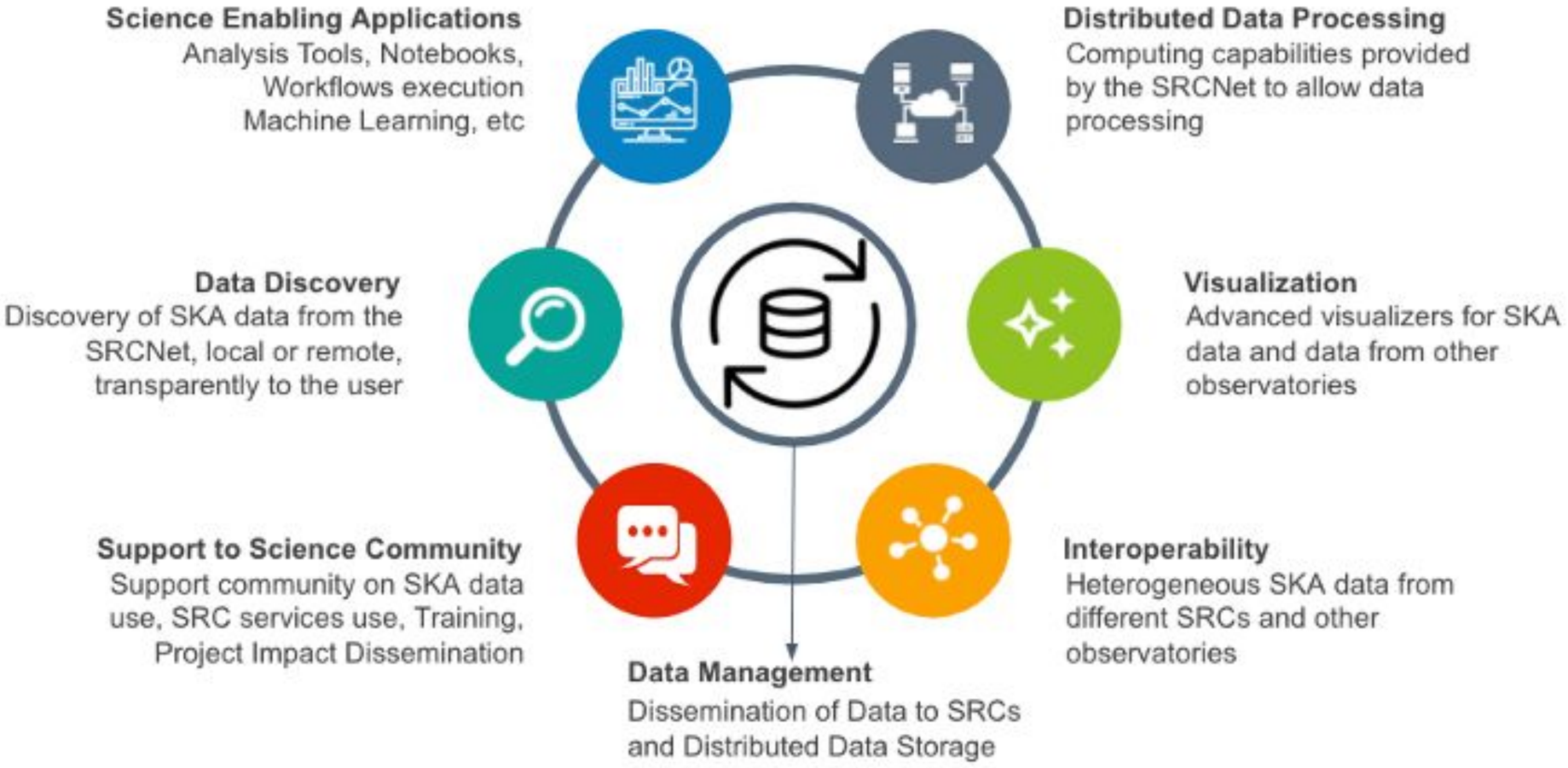


Credit: I. Heywood, SARAO

- Our data are BIG, expecting to deliver ~700 PB/year of *data products*
- Don’t need to be a radio expert to access the SKA!
- Transformational science increasingly relies on multiwavelength data, everyone with great science is welcome :)

SRC Network is critical

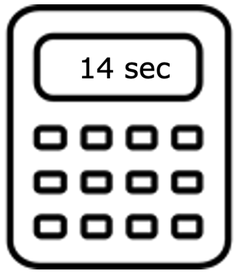
Delivering SKA data products to scientists, storing SKA data for future use, computer facilities to undertake scientific analysis and local user support all fall outside of the construction budget

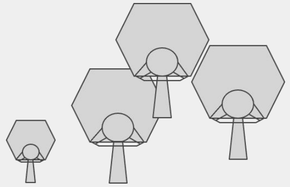


SKAO Operational model (brief summary)

Details in the OEDP: https://www.skao.int/en/resources/402/key-documents
SKAO science data products: A summary document lists many of the kinds of data products we are expecting (details of data formats aren't yet available)

User requests telescope time and data products. SDP is a schedulable part of the telescope (all tools available) + SRCNet requirements





Once taken data may sit in the SDP for up to two weeks before Observation level data products are created and **passed QA**

Once data products have passed QA all raw data will be deleted!

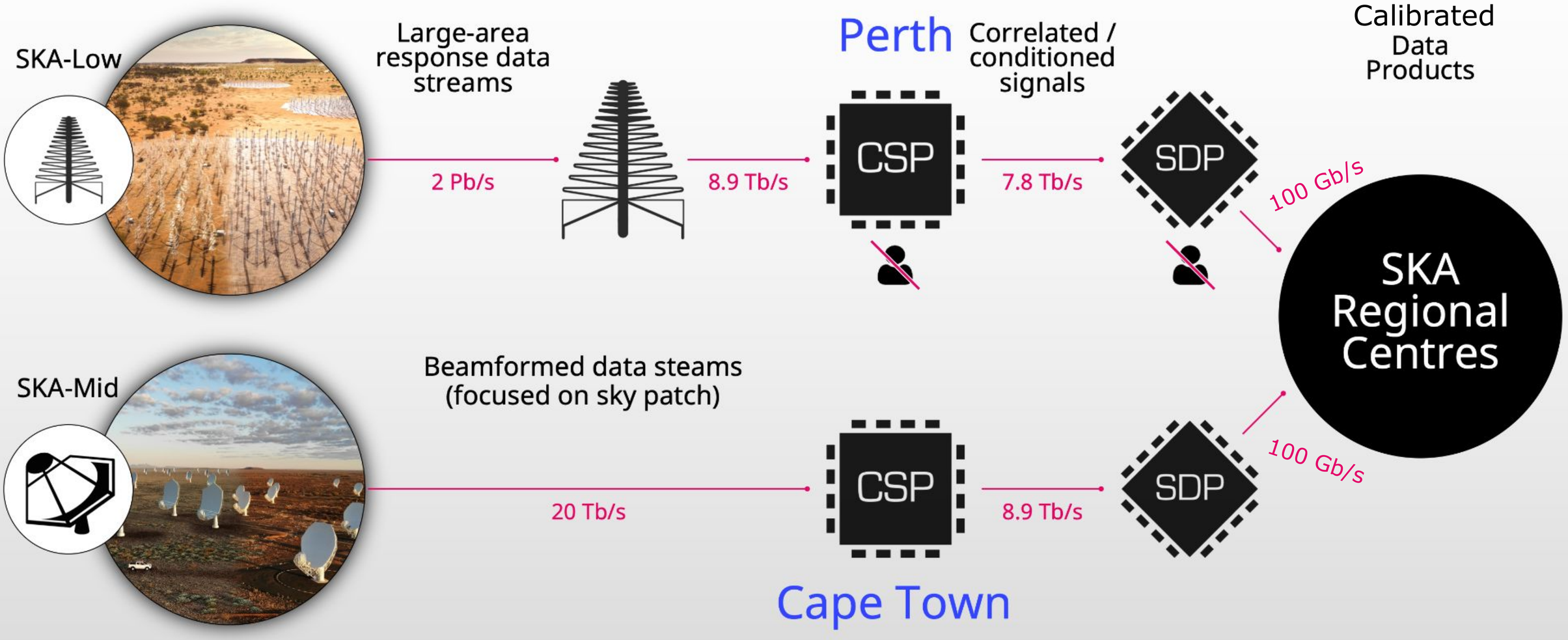
Observation level data products queued for delivery to SRCNet

User requested Project level data products? Created in the SRCNet, responsibility of SKAO

User visualises, analyses, creates Advanced data products on the SRCNet. **WORKFLOW TEMPLATES**

Data made public after the appropriate proprietary period (a year? TBD)

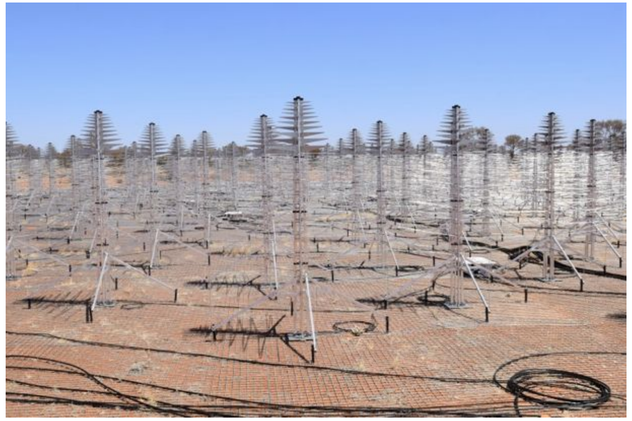
SKA Regional Centres: SKAO data processing stages

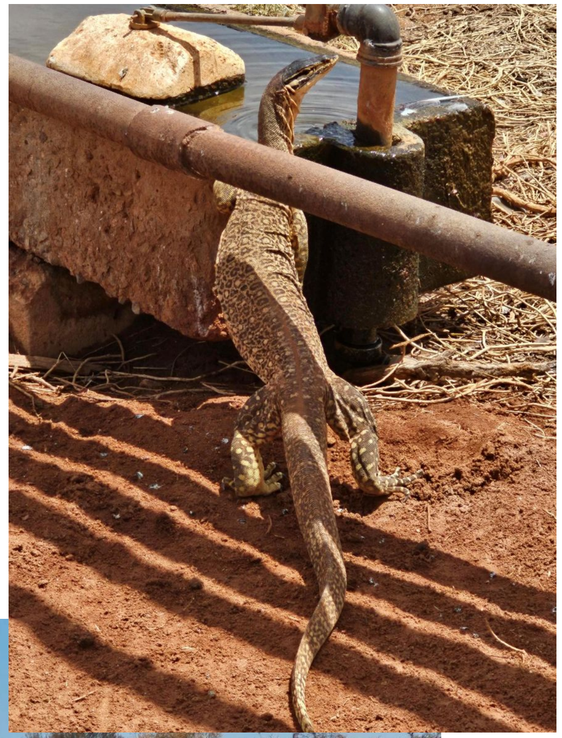


Our construction strategy and timeline

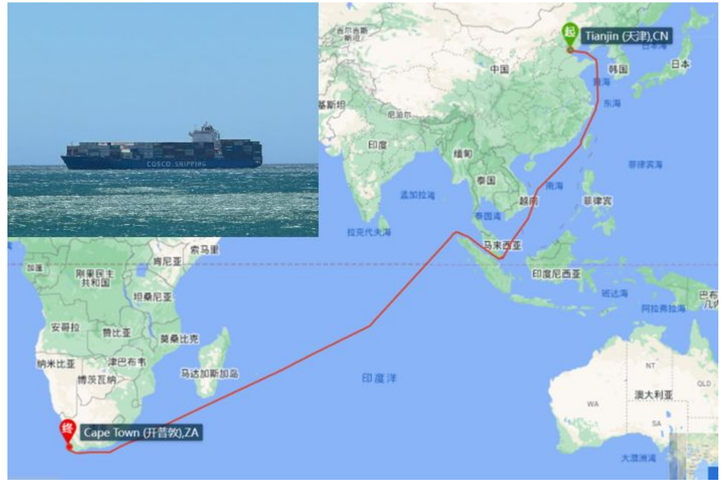
We are now in the construction phase!!!

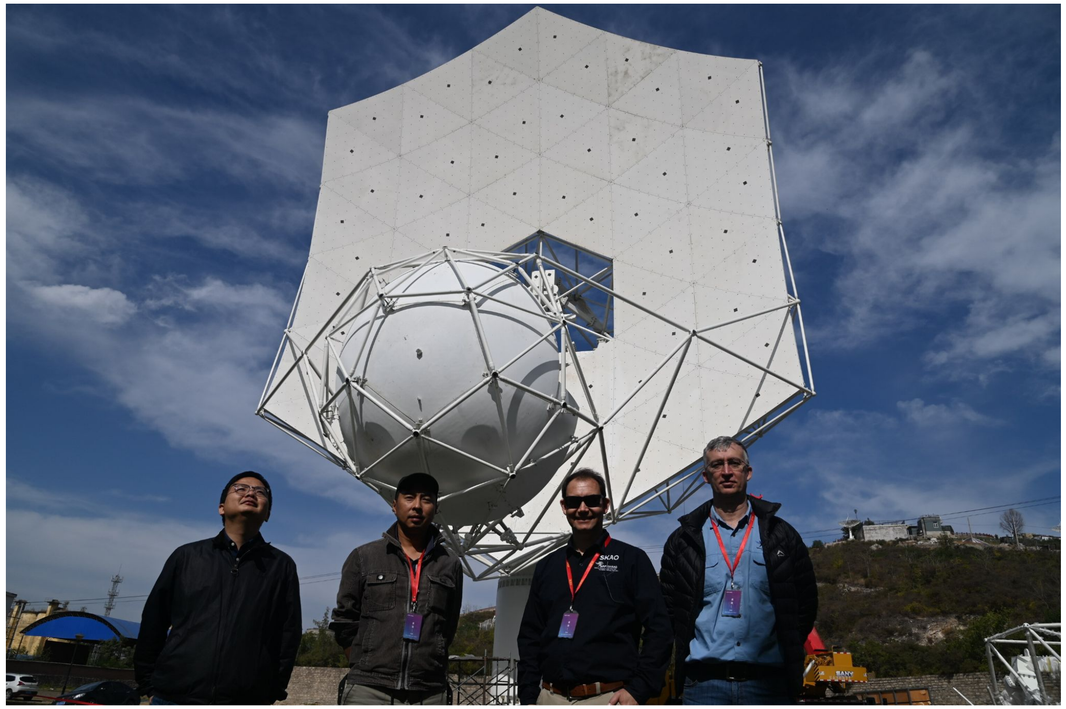
Construction steaming ahead! - Low





Construction steaming ahead! - Mid





Construction Strategy

- **Target**: build the SKA Baseline Design (197 Mid dishes; 512 Low stations: AA4)
- Not all funding yet secured, therefore following Staged Delivery Plan (AA*)
- Develop the earliest possible working demonstration of the architecture and supply chain (AA0.5).
- Then maintain a continuously working and expanding facility that demonstrates the full performance capabilities of the SKA Design.

Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 Apr	2025 Nov
AA2	64 dishes 64 stations	2027 Mar	2026 Oct
AA*	144 dishes 307 stations	2027 Dec	2028 Jan
Operations Readiness Review		2028 Apr	2028 Apr
AA4	197 dishes 512 stations	TBD	TBD

First data release to the community expected in 2026/27 (for science verification)

What does this mean in terms of Operations?

Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 Apr	2025 Nov
AA2	64 dishes 64 stations	2027 Mar	2026 Oct
AA*	144 dishes 307 stations	2027 Dec	2028 Jan
Operations Readiness Review		2028 Apr	2028 Apr
AA4	197 dishes 512 stations	TBD	TBD

Pre science Verification

- Commissioning (+ Assembly, Integration and Verification) primary activity
- SRCs not needed to support AA0.5/AA1 commissioning
- Opportunity for testing (data, transfer, access, pipelines)!

Science Verification

- Data immediately public
- Full dress rehearsal!
- Some SRCNet resources for analysis would be an advantage
- Observed as trickle but also in dedicated blocks
- (+ Commissioning etc ongoing)

Cycle 0

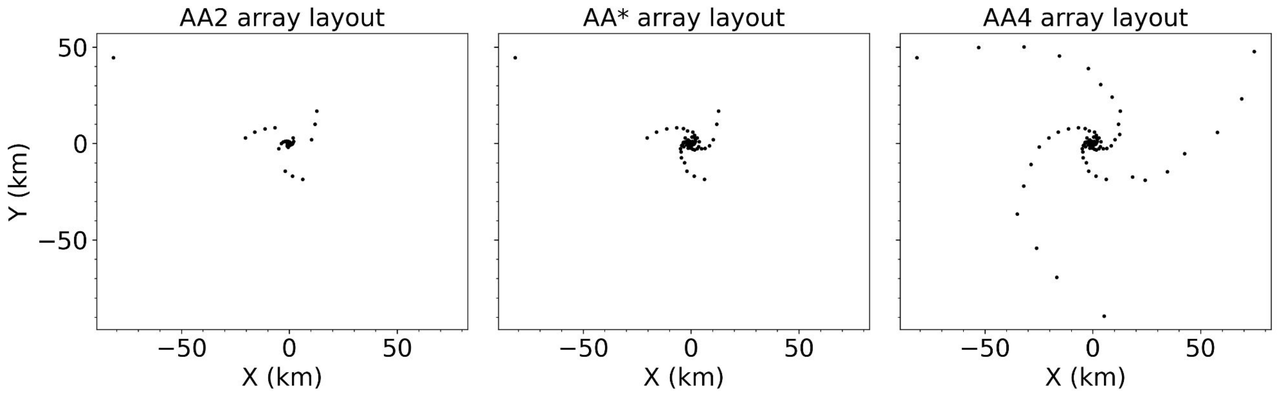
- “Proper” shared risk projects
- Teams, proprietary periods, visualisation, ADP creation etc

Staged delivery: timeline and layouts

Milestone event (earliest)		SKA-Mid (end date)	SKA-Low (end date)
AA0.5	4 dishes 6 stations	2025 May	2024 Nov
AA1	8 dishes 18 stations	2026 Apr	2025 Nov
AA2	64 dishes 64 stations	2027 Mar	2026 Oct
AA*	144 dishes 307 stations	2027 Dec	2028 Jan
Operations Readiness Review		2028 Apr	2028 Apr
AA4	197 dishes 512 stations	TBD	TBD

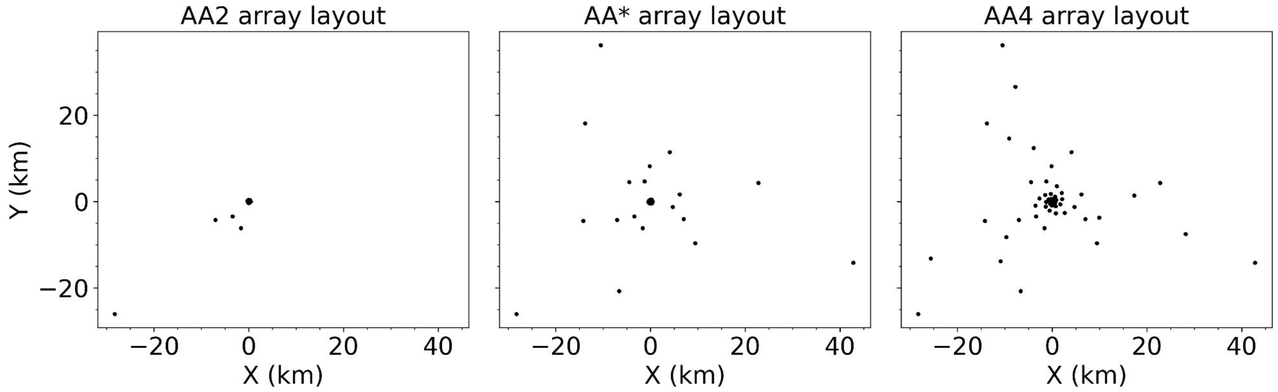
Not just building up in terms of array size, but also capabilities

SKA-Mid



~86 km ~92 km ~152 km

SKA-Low



~34 km ~72 km ~72 km

Predicting our data rates through the stages

Staged delivery: how do the data rates scale?*

***THIS IS WORK IN PROGRESS :)**

- Working assumption that the data flowing out of our Science Data Processor will be 100 Gb/s (i.e. the capacity of the network) at AA4 (baseline design)
- Anticipate less data in the earlier AAs. Estimating those rates is a bit of a challenge..

Data type	Low			Mid		
	AA2	AA*	AA4	AA2	AA*	AA4
	Oct 2026	Jan 2028	TBD	Mar 2027	Dec 2027	TBD
Visibility data - scales with number of baselines	1.6%	36%	100%	11%	54%	100%
Imaging products - scale with the number of pixels which scales with longest baseline	25%	100%	100%	~25%	~25%	100%
Tied array beamformed products - pulsar and transient search, pulsar timing, VLBI scale with the number of beams (not antennas/stations or baseline length)	≤25%	~50%	100%	≤25%	≤50%	100%
Predicted data rate out of SDP (Gb/s)	**5**	**50**	**100**	**10**	**50**	**100**

Can we do better than TBD?

How does this match the data rates/sizes of data products?

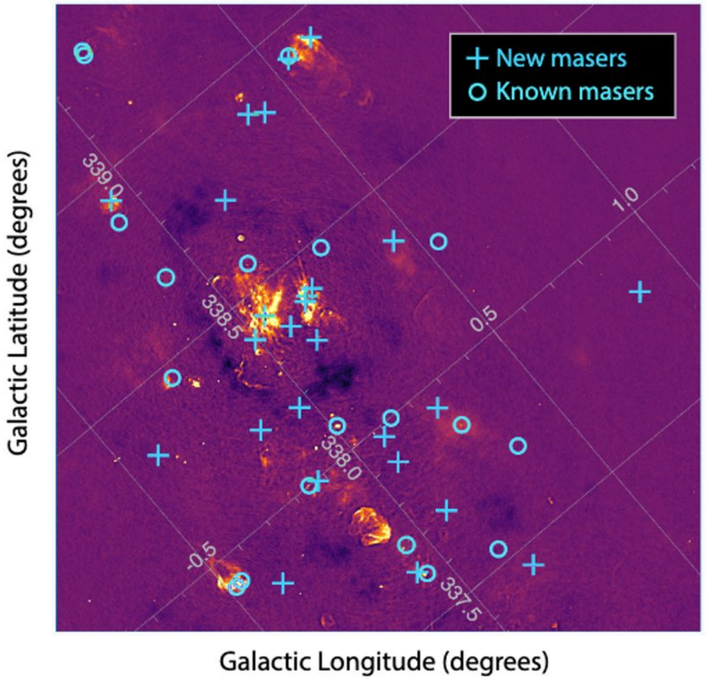
Data product	Low AA2	Low AA*	Mid AA2	Mid AA*
Visibilities - continuum, averaged to 4 sec and 50 kHz	1 Gb/s	20 Gb/s	2.5 Gb/s	11 Gb/s
Visibilities - single zoom, averaged to 4 sec	0.3 Gb/s	7 Gb/s	2.5 Gb/s	11 Gb/s
Continuum image, full field of view, max spatial resolution	0.08 GB	0.3 GB	10 GB	10 GB
Spectral cube 1000 channels, full field of view, max spatial resolution	75 GB	300 GB	10 TB	10 TB
Spectral cube 16000 channels, full FoV, max resolution	12 TB	48 TB	160 TB	160 TB
Pulsar and transient search/ Pulsar timing, each detection, no averaging	15 Gb/s	15 Gb/s	15 Gb/s	15 Gb/s

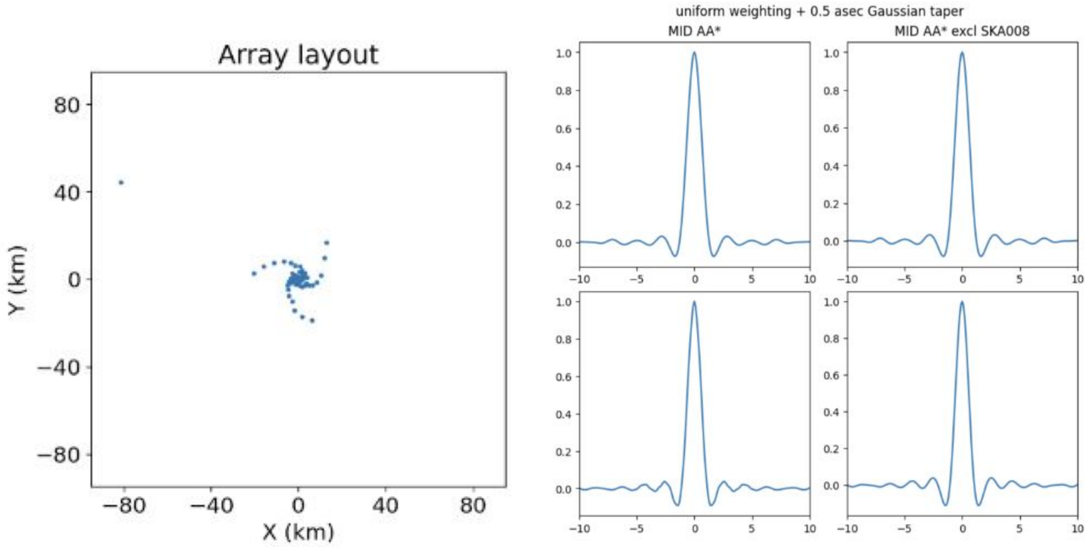
* these numbers are for the full AA* array

- Need to remember that we don't just observe one thing at once, let alone create one data product from each observation
  - Commensality - continuum, zoom,
  - Fast imaging - maybe I am making an image every few seconds, every 10 mins + in a 4-5 hour integration
- The actual data rates are dependent on the science!

What are the challenges?

- The overall data rate depends on
  - How often are we delivering each kind of data product
  - Science needs (what averaging, what image size, how many channels)
  - Commensality (how many things are we doing at once, how many data products)
  - The array used (e.g. how often is the longest Mid baseline useful?)
  - (How efficient we are)
- AA* is a temporary array on the road to AA4
  - How long will we have in AA*?
  - We might imagine we expand from the specified AA* in stages - how and when are antennas added? What if more were added to the spiral arms early? This will increase the data rates!





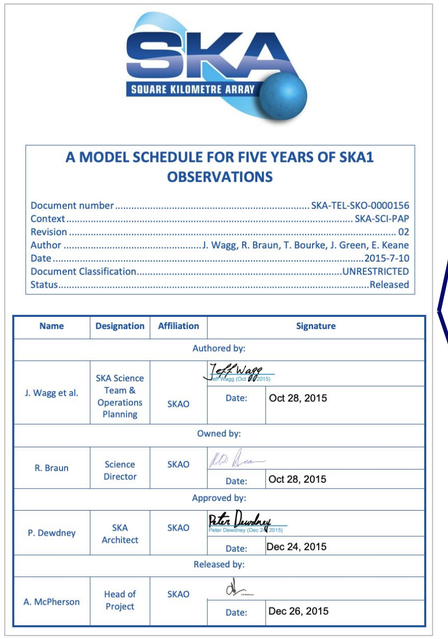
Year in the life of the SKA - the “solution”?

Year in the life of Mid and Low?

- What does a year of Operations look like (“standard” year - cycle 2?)
- Covers as much of the system as we can reasonably include:
  - estimating user support, Helpdesk demand
  - Numbers of proposals
  - The astronomy projects (of all flavours: PI, KSP, ToO + coordinated, joint).
    - Commensality, uniformly reduced data products
  - Data products requested at all stages. What types? What parameters?
    - SDP loads, network links, what is delivered to SRCNet, what resources are needed in SRCNet
  - Weather, RFI, maintenance
  - Power? (how much time does Mid spend slewing, tracking, scanning - how often is it full array?)
  - **Aiming to be representative not perfect, can improve as developments occur**

Why? And why now?

- We now know enough detail about systems, science plans etc to develop a quite sophisticated and accurate representation of what a year might look like in full operations
  - Important for operational planning, development of systems
    - User support models
    - Are we staffed well? Do we have enough operators, Astronomer on Duty etc?
    - Proposal systems
    - Mode/data product usage
    - SDP planning (and conversely how we schedule observations so that SDP can manage resource heavy projects).
    - Commensality planning
    - VLBI plans
    - Network capacity
    - …
    - -> developing a realistic project schedule underpins much of this
  - User expectations



SQUARE KILOMETRE ARRAY

A MODEL SCHEDULE FOR FIVE YEARS OF SKA1 OBSERVATIONS

Document number SKA-TEL-SKO-0000156
Context ..... SKA-SCI-PAP
Revision 02
Author ..... J. Wagg, R. Braun, T. Bourke, J. Green, E. Keane
Date 2015-7-10
Document Classification ..... UNRESTRICTED
Status ..... Released

Name	Designation	Affiliation	Signature	
Authored by:				
J. Wagg et al.	SKA Science Team & Operations Planning	SKAO		Date: Oct 28, 2015
Owned by:				
R. Braun	Science Director	SKAO		Date: Oct 28, 2015
Approved by:				
P. Dewdney	SKA Architect	SKAO		Date: Dec 24, 2015
Released by:				
A. McPherson	Head of Project	SKAO		Date: Dec 26, 2015

SKAO

A Year in the Life of SKA-Low

SKA-TEL-SW-0000000	Revision 01
Classification:	UNRESTRICTED
Document type:	DTE
Date:	2023-02-??
Status:	DRAFT

SKAO

A year in the life of SKA-Mid

SKAO-TEL-0000000	Revision 01B
Classification:	UNRESTRICTED
Document type:	DTE
Date:	2021-06-22
Status:	DRAFT

The project list

Developed with our Science Working Groups and associated documentation

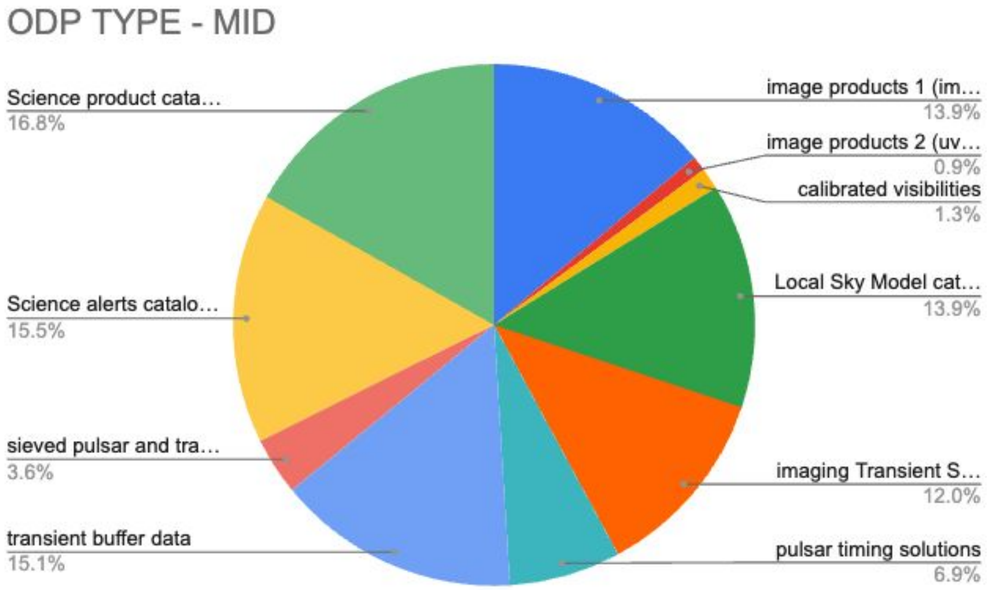
Ref.	Description	Allocated Time	Project Type	Primary science driver(s)	Primary observing mode(s)	Primary telescope tracking mode(s)	Primary ODP type(s)	Primary CBF Output Data Rate (Gb/s)	FSPs primary
LOW-KSP-A	EoR tomography / power spectrum deep	438	KSP - STD	EoR	CONT	SIDEREAL, DRIFT	A,C,D,J	1676	24
LOW-KSP-B	Legacy shallow	438	KSP - STD	Exgal Cont, Exgal Spec, Mag, HiGal, Cos, OG	CONT	SIDEREAL, DRIFT	A,D,J	984	24
LOW-KSP-C	Legacy medium	438	KSP - STD	Exgal Cont, Exgal Spec, Mag, HiGal, Cos, OG	CONT	SIDEREAL, DRIFT	A,D,J	984	24
LOW-KSP-D	Legacy deep	438	KSP - STD	Exgal Cont, Exgal Spec, Mag, HiGal,	CONT	SIDEREAL, DRIFT	A,B,D,J	984	24
LOW-KSP-E	Galaxy clusters/relics	438	KSP - STD	Exgal Cont, Mag, Cos	CONT	SIDEREAL	A,D,J	705	14
LOW-KSP-F	CMI exoplanets / low-mass stars	438	KSP - STD	CdL, OG, Tr, Mag	CONT, FLOWTHROUGH, TRANSIENT MODE 2?	SIDEREAL	A,D,E,G,I,J	185	32
LOW-KSP-G*	Solar monitoring	438	KSP - STD	SHI	CONT, DYNAMIC	NON	A,D,E,I,J,K	3648	32

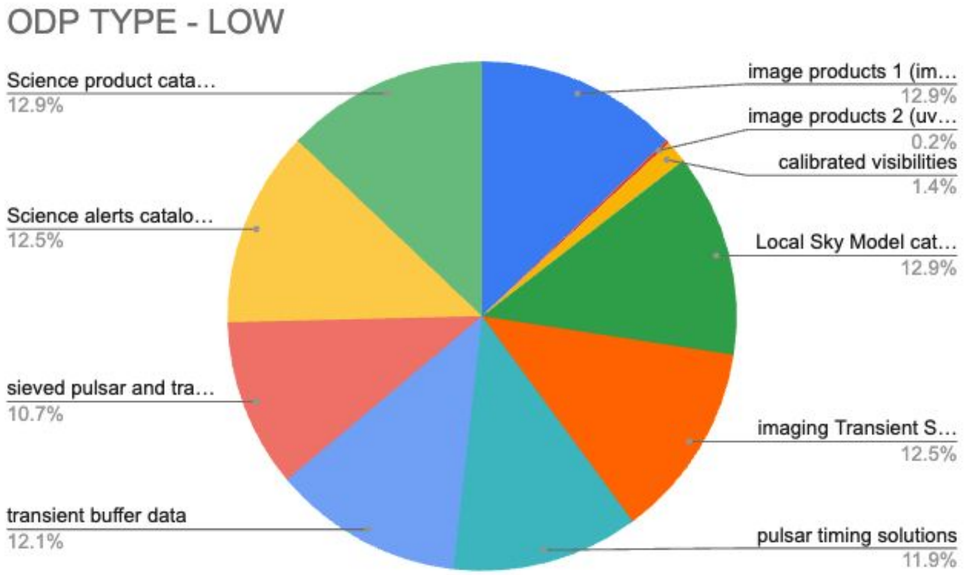
Work in progress

- Processing load on SDP computed using a parametric model that has recently been updated
  - "Primary project"
  - Commensal projects
- Output data product sizes based on scientifically representative parameters

The project list: data product relative delivery rates

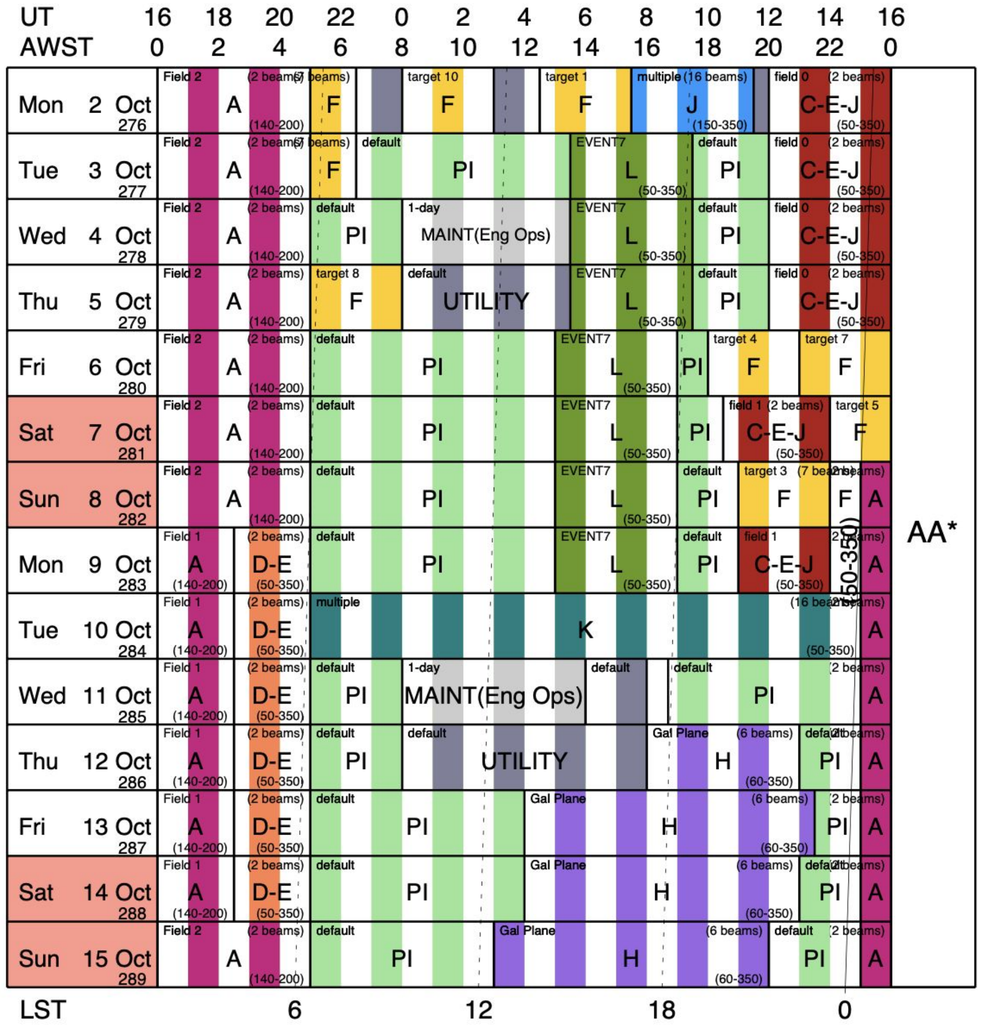
*these are currently not for all projects/full cycle

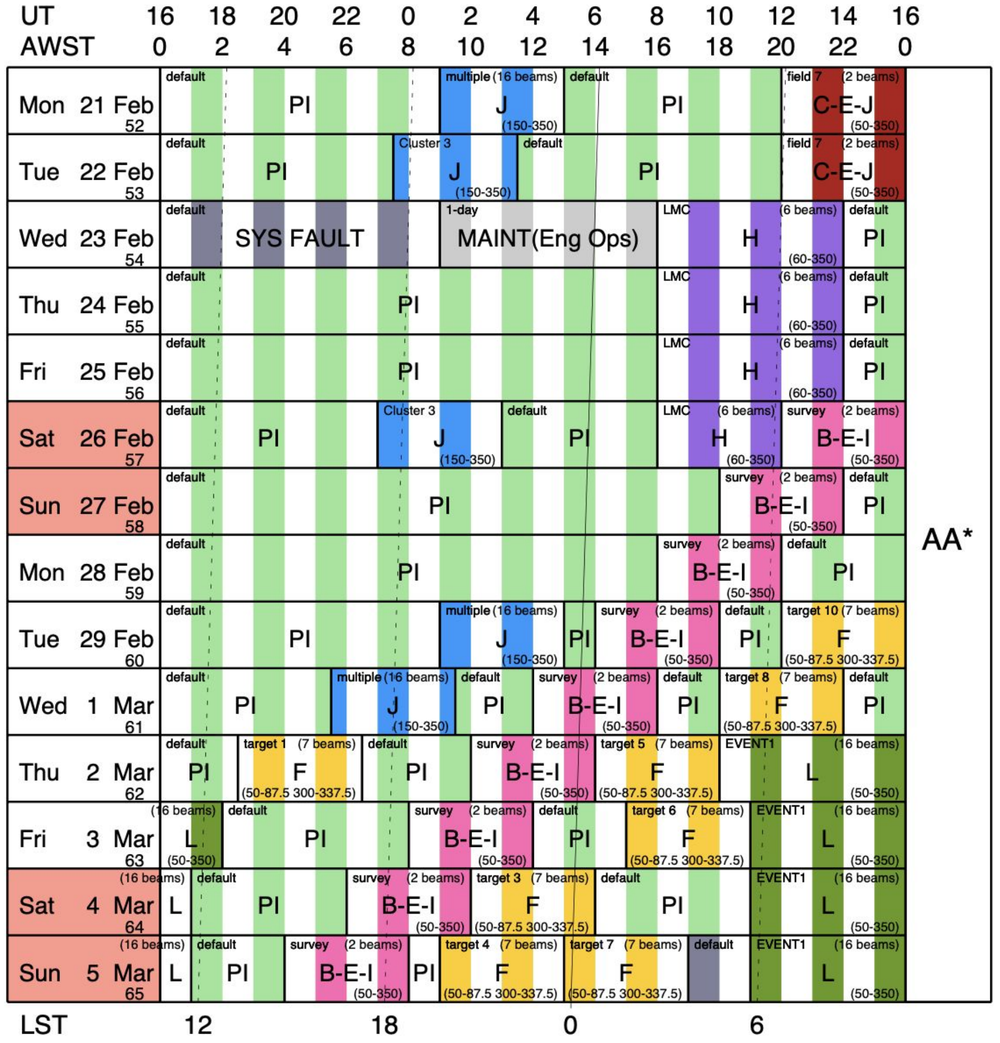




The project list: create a cycle-long schedule!

This will allow us to predict the data flow through the system and allow us to determine where the bottlenecks are (if there are any) and also provide a better estimate of requirements in AA* (and AA4 in the future)





Summary

- SKA is now in construction
- We are rolling out the telescopes in stages
- Our output data rates will grow through these stages, building to 100 Gb/s
- Exactly how these data rates evolve is not just based on array size but also science capability and needs
- Data rates through the AAs can be estimated by scaling from our baseline design (but this also had assumptions)
- We are currently undertaking a huge piece of work to simulate what a year in the life of the SKA might look like
  - Will tell us all kinds of things, but in this context, importantly the flow of data through each stage of the system (2 weeks of “busy”time starting next week so expecting lots of progress towards something that can be shared!)

Thank you!

Shari.Breen@skao.int

We recognise and acknowledge the Indigenous peoples and cultures that have traditionally lived on the lands on which our facilities are located.

SKAO

www.skao.int